

### **Amendments to the Specification**

Please replace paragraph [018], page 5, with the following:

[018] Figure 3 is a dynamic Bayesian network representation of the model of ~~Figure 2A~~ Figure 2; and

Please replace paragraph [020], page 5, with the following:

[020] As shown in Figure 1, our invention uses unsupervised learning to determine a structure of a video 101, i.e., an audio-visual stream. Knowing the structure, we can detect events in the video that are consistent with the structure. The structure is in the form of a hierarchical statistical model 200, see ~~Figures 2A-2B~~ Figures 2-3. The hierarchical model can be based on hierarchical hidden Markov models (HHMM) or hierarchical Gaussian mixture models (GHMM).